DOCUMENT RESUME

ED 319 789 TM 015 044

AUTHOR Marsh, Herbert W.; And Others
TITLE Self-Concepts of Young Children Aged 5 to 8: Their Measurement and Multidimensional Structure.
PUB DATE 20 Mar 90
NOTE 44p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Child Development; Comparative Testing; *Factor Structure; Foreign Countries; Goodness of Fit; Grade 1; Grade 2; *Interviews; Kindergarten Children; *Multidimensional Scaling; Primary Education; *Self Concept Measures; *Young Children
IDENTIFIERS Australia; Confirmatory Factor Analysis; *Self Description Questionnaire

ABSTRACT

The purposes of the present investigation are to evaluate a new, adaptive procedure for assessing multiple dimensions of self-concept for children younger than 8 years and to examine related theoretical issues. The multidimensional, hierarchical structure of self-concept is well established for older children, but there is a paucity of research and appropriate instruments for very young children. A limited amount of research suggests that self-concept is poorly differentiated and that a general self-concept may not even exist. A sample of 501 students in kindergarten and first and second grades in Sydney (Australia) completed a variation of the Self-Description Questionnaire I (SDQI) using a new individual interview technique. At each grade level, confirmatory factor analyses identified all eight SDQI scales, including the general self-concept scale. With increasing age, the fit of the eight-factor model improved, and the size of correlations among the eight SDQI scales decreased, implying that self-concept becomes more differentiated with age. The results demonstrate that appropriately measured self-concepts are much better differentiated for very young children than previously assumed. A 55-item list of references and seven data tables are included. (Author/TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED319789

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

HERBERT W. MARSH

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Self-concepts Of Young Children Aged 5 to 8: Their Measurement and Multidimensional Structure

Herbert W. Marsh, Rhonda Craven, and Raymond Debus

University of Sydney, Australia

20 March, 1990

Running Head: Self-concepts of Young Children

TM 015 044

BEST COPY AVAILABLE

Self-concepts Of Young Children Aged 5 to 8:
Their Measurement and Multidimensional Structure

ABSTRACT

The purposes of the present investigation are to evaluate a new, adaptive procedure for assessing multiple dimensions of self-concept for children younger than 8 and to examine related theoretical issues. The multidimensional, hierarchical structure of self-concept is now well established for older children but there is a paucity of research and appropriate instruments for very young children. A limited amount of research suggests that self-concept is poorly differentiated and that a general self-concept may not even exist. 501 students in kindergarten, 1st, and 2nd grades completed a variation of the SDQI using a new individual interview technique. At each grade level confirmatory factor analyses identified all 8 SDQI scales — including the General self-concept scale. With increasing age the fit of the 8-factor model improved and the size of correlations among the 8 SDQI scales decreased, implying that self-concept becomes more differentiated with age. The results demonstrate that appropriately measured self-concepts are much better differentiated for very young children than previously assumed.

A positive self-concept is valued as a desirable outcome and as a potential mediating influence leading to other desired outcomes such as academic achievement. Despite the thousands of self-concept studies conducted with older students, there has been little research conducted with children below the age of 10. This is unfortunate as this developmental period may be critical in the formation of a positive self-concept — particularly in educational settings. This lack of research stems apparently from the dearth of instruments appropriate for measuring self-concepts for children of this young age.

Prior to the 1980s reviews of self-concept research based on responses by older children and young adults noted a lack of theoretical models and appropriate measurement instruments. The Shavelson, et al. (1976) model of a self-concept was valuable in remedying some of these problems. They proposed self-concept to be a multifaceted, hierarchical construct that became increasingly distinct with age. In their model a general facet at the apex of the self-concept hierarchy is divided into academic and nonacademic components of self-concept. The nonacademic component is divided into physical, social, and emotional components, whereas the academic component is divided into academic facets in particular subject areas such as English and mathematics. Particularly during the last decade, as researchers have developed apparently better self-concept instruments based on stronger theoretical models, support for the multidimensionality of self-concept for older children and young adults has become well established (e.g., Byrne, 1984; Dusek & Flaherty, 1981; Fleming & Courtney, 1984; Harter, 1982; Marsh, Byrne & Shavelson, 1988; Marsh, in press-a).

Harter (1983, 1985, 1986) evaluated alternative models of self. Like Shavelson, et al. (1976), she rejected models depicting self-concept as a simple summated score based on items reflecting different domains (e.g., Coopersmith, 1967) and argued for a multidimensional perspective that recognizes specific domains such as the physical, social, and academic facets of self. She also argued for the usefulness of a relatively unidimensional, global self-concept like that described by Rosenberg (1979). In discussing the Rosenberg approach to global self-worth she noted that it "finesses the complexities of the underlying hierarchy of discrete judgments that may be responsible for such an overall judgment about the self" (Harter, 1986, p. 141). Combining these two perspectives, Harter recommended a hybrid approach that includes both content specific scales and a global scale.

Empirical support for a multidimensional, hierarchical self-concept proposed by Shavelson et al. (1976) and for the hybrid approach proposed by

Harter is particularly strong in research using the Self Description Questionnaire (SDQ) instruments (see Marsh, 1989, in press-a, for an overview). Three SDQ instruments designed for children of differing ages are based on the Shavelson et. al. model. In research using the SDQI with young children, limitations in children's ability to respond to questionnaires has been overcome in part by reading aloud the SDQI items. Using this approach, Marsh, Barnes, Cairns and Tidman (1984) tested large samples of students in grades 2-5. Even for the second grade children, the SDQ factor structure was reasonably well defined and the internal consistency estimates for the SDQI scales were moderate. In subsequent confirmatory factor analyses (CFAs) using the LISREL approach to testing factorial invariance, Marsh and Hocevar (1985) found that the SDQI factor structure (factor loadings) were reasonably invariant across the different ages. Consistent with the Shavelson et al. (1976) hypothesis that self-concept becomes more differentiated with age, they found that correlations among the factors became smaller with age. Whereas a similar pattern of hierarchical factors was identified at each age, the self-concept factors became more distinct and the hierarchy became weaker with age.

The purposes of the present investigation are to evaluate a new, adaptive procedure for assessing multiple dimensions of self-concept for children younger than 8 using a variation of the SDQI and to examine theoretical issues related to the factorial structure (or dimensionality) of self-concept for these young children. The theoretical basis of this research derives largely from Harter's research (1983, 1985, 1986; Harter & Pike, 1984; Silon & Harter, 1985) that focuses specifically on the self-concepts of very young children, the Shavelson, et al. (1976; Marsh & Shavelson, 1985) model of self-concept, and previous SDQ research (e.g., Marsh, 1988; Marsh, in press-a; Marsh, Byrne & Shavelson, 1988). The development of appropriate measurement devices for children of this age has important theoretical and practical implications. Theoretically, the study provides important evidence about the ability of very young children to differentiate specific facets of self-concept and to form a generalized conception of self, and about age and gender differences in self-concepts for very young children. From a practical perspective, the ability to measure the self-concepts of very young children provides an important outcome measure for for teachers to better understand their students and for a wide variety of interventions designed for young children.

The Self-concepts of Very Young Children

Despite a growing consensus in findings for older children, the status of the multidimensionality and the factorial structure (or dimensionality)

of self-concept for very young children is unclear. The extent to which self-concept is differentiated for young children apparently reflects the cognitive development of the child (Stipek & McIver, 1989; Silon & Harter, 1985) and the appropriateness of the instrument used to assess self-concept (Harter & Pike, 1984; Stipek & MacIver, 1989). Whereas most researchers assume that the dimensions of self-concept become increasingly distinct with age, there is limited evidence on the the distinctiveness of self-concept factors at different ages (Marsh, 1989). Harter and Pike (1984; Harter, 1983, 1986), for example, suggested that for children younger than 8 a global sense of self-worth does not exist and specific facets of self-concept are not well differentiated. Stipek and MacIver (1989) noted that very young children have a poorly differentiated concept of academic competence but that it becomes better differentiated from other facets (e.g., social competence) during elementary school years. They also suggested, however, that the lack of differentiation may reflect problems with existing measurement instruments and recommended the use of more appropriate assessment procedures. Wylie (in press) also noted the difficulties in assessing self-concept in young children, stating that "it is generally conceded that preschool children have some descriptive and evaluative self-conceptions, but the problem of measuring them is obviously a thorny one" (p. 70). Perhaps, as appears to have been the case for research with older children, progress in theory and research for very young children will be stimulated by the development of better multidimensional measurement instruments.

In this section we discuss three issues related to evaluating the self-concepts of very young children that are the basis of the present investigation. First, there is a need to evaluate how to measure self-concept most effectively for very young children. As proposed by Harter (1983, 1986; Harter & Pike, 1984) this may require simplified item contents or pictorial representations, simplified response formats, and individually based interviews instead of conventional paper-and-pencil tests that are group administered. Second, there is a need to evaluate the factorial structure of self-concept responses for young children and to determine whether factors like those identified in the responses of older children can be found (e.g., Harter, 1982; Marsh, 1988, in press-a). Third, there are important, unresolved issues surrounding the status of general self-concept in very young children. For example, Harter and Pike (1984) claim that general self-concept evolves from the integration of domain-specific facets of self, and thus does not exist prior to the age of 8.

The Measurement of Self-concept For Young Children

Wylie (in press) has recently updated her classic review of self-concept measures (Wylie, 1974; 1979), and was particularly concerned to identify suitable instruments designed for very young children. She selected what were apparently the two best of existing instruments. She noted, however, that there was insufficient evidence to adequately evaluate either instrument, but included them because no more fully developed self-concept measures were available for very young children. One of these instruments relied on items from a variety of different domains to infer a global, undifferentiated self-concept. Because an important focus of the present investigation is to determine the factorial structure of self-concept and whether young children are able to differentiate among specific facets of self, this instrument is of limited relevance.

The second instrument considered by Wylie, the Harter and Pike (1984) instrument, was designed to measure four self-concept scales: physical, cognitive, peers, and maternal. Each scale is defined by six bipolar items that are represented by parallel verbal statements and pictures. For example, respondents are shown two pictures, one in which the target child appears with one other child and one in which the target child appears with five other children. The respondent is told that the first target child doesn't have many friends to play with and that the other target child has lots of friends to play with. The respondent's first task is to indicate which target child is most like the respondent. After choosing a target child, the respondents indicate whether they are a lot like the chosen target or just a little bit like the target child. For each item there are more specific prompts such as do you have "a whole lot of friends to play with" or "pretty many" (see Harter & Pike, 1981; 1984). This two-stage response format consisting of two dichotomous responses results in a 4-point response scale.

Harter and Pike (1984) emphasized a number of important features of this scale including: (a) items appropriate to the developmental level of the children, (b) the pictorial format, and (c) the 4-point response scale that provides for a greater range of responses than typical dichotomous responses. Although not emphasized by Harter and Pike, another potentially important feature is that they administered their instrument individually to each child instead of using a group administration procedure as is typical with older children. Whereas they did not provide details of their individual administration procedures, this procedure offers the possibility for the administrator to ensure that the child understands an item and, if necessary, to clarify the meaning of an item for the child and the child's

response. Of particular relevance to the present investigation, no global self scale was included because Harter's theoretical and empirical research indicated that global self-concept does not evolve until approximately the age of 8 (Harter & Pike, 1984).

The Factorial Structure of Self-concept in Responses By Young Children

Predictions about how self-concept and its factorial structure evolves with age have been proposed from a variety of theoretical perspectives. A basic assumption of the Shavelson model is that self-concept becomes more differentiated with age. Markus and Wurf (1987) note that the structure of self depends on both the information available to an individual and the cognitive ability to process this information. Harter (1983, 1985) proposed a model in which self-concept becomes increasingly abstract with age. Her review of previous research suggested that self-conceptions shift from concrete descriptions of behavior in early childhood, to trait-like psychological constructs (e.g., popular, smart, good looking) in middle childhood, to more abstract constructs during adolescence.

Harter (1983, 1985), consistent with her proposal that self-concept becomes increasingly integrated with age, reported that the concept of general self-worth does not evolve before the age of about 8. Harter and Pike (1984) reported that below age 8 children do not understand general self-worth items or provide unreliable responses. Subsequent research (Silon & Harter, 1985) suggested that mental age may be more important than chronological age. Harter's assumption is apparently in direct contradiction Coopersmith's (1967) contention that distinctions among specific domains are made by young children "within the context of the over-all, general appraisal of worthiness that children have already made" (p. 6). Interestingly, interpretations by each researcher are based on the failure of factor analyses to identify the intended self-concept scales in their respective instruments, although the failure to support an a priori factor structure is a weak basis of support. Whereas Harter (1983; also see Marsh & Richards, 1988, Marsh & Smith (1982), Shavelson, et al., 1976) called into question the logical basis of Coopersmith's conclusion, there does not appear to be compelling support for either of these opposing perspectives.

Specific factors. The Harter and Pike (1984) instrument was designed to measure four self-concept scales. Factor analyses, however, provided support for only two factors: competence (incorporating the physical and cognitive scales) and social acceptance (incorporating the peer and maternal scales). The authors noted that the factor structure is less differentiated than typically found for older children, thus supporting the frequently noted

assumption that the structure becomes more differentiated with age. Based primarily on support for this two-factor model, Harter and Pike suggested that young children do not differentiate among competencies in different areas although they do differentiate between general competence and general social acceptance. Support for a similar two-factor model was also found in the Silon and Harter (1985) study of responses by children with chronological ages of 9-12 who had mental ages of less than 8.

Harter and Pike's (1984) failure to support their a priori four-factor structure provides a weak basis of inference about the structure of self-concept, particularly when analyses were based on exploratory factor analyses rather than the methodologically stronger CFA that allows the researcher to specify the model to be tested (e.g., Marsh & Hocevar, 1985). The failure to separate even the physical and academic components that are so robust in responses by slightly older students is surprising and requires further scrutiny. Correlations among the physical and academic scales in the Harter and Pike study varied from .43 to .56 and did not approach 1.0 even after correction for unreliability. Furthermore, Harter and Pike (1984) noted that when the four scales were correlated with external criteria (e.g., teacher ratings, choice behavior, being held back a grade) there was support for the separation of the physical and academic scales. Also, other facets — particularly physical appearance — were not included that could, perhaps, be differentiated from other areas of self, are appropriate for very young children, and do not fit easily into the categories of either competence or social acceptance. Finally, even support for their conclusion that self-concept becomes more differentiated with age was weak. They noted that their factor structure was less differentiated than that found with older children, but they did not administer their instrument to older children. More importantly, they did not offer any evidence that self-concept became more differentiated with age for the 4-7 age range that they considered, suggesting instead that the factor structure was similar for their preschool/kindergarten sample and for their 1st/2nd grade sample. In summary, whereas the Harter and Pike (1984) study is important, it may be premature to conclude that children can only differentiate two broad components of self.

Global self-concept. Harter and Pike (1984) did not include a global self scale on their instrument because "both theory (see Harter, 1983) and empirical findings have led to the conclusion that children are not capable of making judgments about their worth as a person until approximately the age of 8" (Harter & Pike, 1984, p. 1970). In discussing the Harter and Pike results, Silon and Harter (1985) noted that "below the age of 8, children

either do not understand the [general] self-worth items, produce extremely unreliable estimates, or both" (p. 227). However, neither the theoretical nor the empirical support for this conclusion is presented in sufficient detail in either of these studies to be evaluated adequately.

In Harter and Pike (1984), as elsewhere (e.g., Harter, 1986; Silon & Harter, 1985), the reader is referred to Harter (1983) for the theoretical basis of this conclusion. Whereas Harter (1983) provided a rich theoretical framework for the development of self, she did not directly address the theoretical rationale for the assertion that general self-concept does not exist before the age of 8 and her general theoretical discussion did not provide a firm basis for it. For example, Harter (1983) noted that the capacity for limited hierarchical organization first appears during the concrete operational period (e.g., I'm smart because I'm good at reading, spelling, and mathematics), but she also noted that when general trait labels first become available, children tend to use an "all-or-none" thinking. Whereas children are unable to integrate specific components of self to form a global self-concept at Stage I in Harter's (1983) developmental schema, children at this stage think they are all good (or all bad) across a wide variety of domains. In the Harter schema, however, it is not until stage IV (middle adolescence) that children are capable of higher-order abstractions. From this perspective is not clear why children aged 8-12 are able to have general self-concepts, though her empirical research indicates that they are. Elsewhere Harter (1983) noted that whereas very young children can experience emotions such as pride and shame, definitions of these emotions focused on others (Dad was proud of me because ...) at ages 5-7 and that self-references did not emerge until the age of 8. Whereas it is reasonable that very young children do not have the cognitive capacity to integrate specific components of self to form a general sense of self, this presupposes that general self follows rather than precedes self-concepts in specific domains. If self-concepts in specific domains are derived from a general sense of self, then children may not require this integrative capacity in order to experience an overall sense of self. On the basis of Harter's (1983) review it appears likely that the processes underlying the formation of general self vary systematically with age but it appears premature to conclude that very young children do not have a general self-concept.

Harter and Pike (1984) provided even less discussion of the empirical basis for their assumption. Silon and Harter (1985), referring to the Harter and Pike research, noted that the general self scale lacks reliability, but

no reliability estimates were presented for the global scale in either study. Silon and Harter (1985) administered Harter's (1982) instrument designed for older children to children who had chronological ages of 9-12 but mental ages less than 8. They reported that items from the global scale did not emerge as a separate factor and did not load consistently on either of the two factors (general competence and acceptance) that did emerge. Responses by these older but intellectually disadvantaged children, however, may provide a dubious basis of generalization to responses by very young children. Furthermore, here again it must be noted that the failure to replicate a factor structure is a weak basis of support for this contention. A much stronger test would have been possible had the researchers used CFA as in Marsh and Hocevar (1985) instead of exploratory factor analysis.

The Present Investigation

In the present investigation 501 young children in kindergarten, 1st, and 2nd grades responded to an individually presented version of the SDQI using a modified response format. Subsequently, about 2 weeks after the individually administered SDQIs were collected, the typical group administered SDQI was collected from 1st and 2nd grade children. The factor structure and responses for the group and individually administered versions were compared. The major aims of the study are: (a) to establish the psychometric properties of these responses and to determine whether the SDQI factor structure found with older children can be replicated; (b) to test Harter's claim that general self-concept does not exist before the age of 8; (c) to test the Shavelson et al. (1976) hypothesis that self-concept becomes more differentiated with age and provide more specific data on how the factor structure of self-concept varies in the age range of 5 to 8; (d) to compare individual and group administered versions of the instrument, and (e) to evaluate sex and age differences for these very young children.

Methods

Subjects

A total of 501 students from kindergarten (n=163), grade one (n=169), and grade 2 (n=169) participated in the study. Children in each of the three grade levels were predominantly 5 years of age (kindergarten), 6 years of age (1st grade), and 7 years of age (2nd grade). The subjects came primarily from middle class families and attended one of three infant schools in suburban metropolitan Sydney, Australia.

Instruments: The SDQI

The SDQI (Marsh, 1988, in press-a) is is one of a set of three instruments designed to measure multiple dimensions of self-concept for preadolescents (SDQI), for early and middle adolescents (SDQII), and for

late adolescents and young adults (SDQIII). The instruments are based on the Shavelson et al. (1976) model of self-concept and the facets of self proposed in that model. More than 30 published factor analyses have identified the factors that each instrument is designed to measure. Other research has shown that: (a) the reliability of each scale is generally in the 0.80s and 0.90s whereas correlations among the factors are quite small (median rs less than 0.20), (b) the self-concept responses are substantially correlated with self-concepts in matching areas inferred by significant others, (c) academic achievement indicators are substantially correlated with academic areas of self-concept but nearly uncorrelated or even negatively correlated with nonacademic areas of self-concept and general self-concept, (d) self-concept factors are systematically and logically related to a variety of other constructs including age, gender, locus of control, self-attributions for the causes of academic successes and failures, physical fitness and participation in sports, and interventions designed to enhance self-concept. This research provides strong support for the construct validity of responses to the SDQ instruments for children aged 10 or older, and perhaps children as young as 8. In her review of academic self-concept measures, Byrne (in press) concluded: "it is apparent that the SDQ I, II and III, without question, are the most rigorously validated instruments currently available to researchers; each is solidly linked to the hierarchical model of SC [self-concept] theory, and their separate ACS [academic self-concept] subscales have been shown to be psychometrically sound" (p. 29).

The SDQI (Marsh, 1988) assesses three areas of academic self-concept (reading, mathematics, and school self-concept), four areas of non-academic self-concept (physical ability, physical appearance, peer, and parent relations) and a general-self scale. Three total scores can also be measured on the basis of these scales; academic self-concept (the average of reading, mathematics, and school self-concept scales), non-academic self-concept (the average of physical, appearance, peer, and parent relations self-concept scales), and total self (the average of academic and non-academic total scales). For both the standard (group administered) and individually administered versions of the SDQI each of the 8 SDQI scales was defined by responses to the same 8 positively worded items. On the standard SDQI there are an additional 12 negatively worded items. Because previous research has shown that children have trouble responding appropriately to the negatively worded items (Marsh, 1986a), they are not included in the scores derived from the SDQI (Marsh, 1988). For purposes of the individually administered

SDQI used here, the negatively worded items were excluded altogether. As described below, the response scale typically used with the SDQI was also altered for purposes of just the individually administered responses.

Procedures

Procedures for the administration of the standard SDQ (see Marsh, 1988) were adjusted to enable the modified SDQ to be administered by trained interviewers to subjects in an individual interview. The interviewers were 120 college students in a primary teacher education program who already had experience working with young children. Interviewers were from a large teacher education course in social studies that contained a component on self-concept research. Participation consisted of attending a preliminary training session and subsequently testing children from each of the three age groups. Training consisted of attending a two hour training session in which procedures for administering the instrument were explained, a ten minute training video of a kindergarten child responding to all procedures was viewed, and a ten minute administration practice session took place with another trainee interviewer responding to the questionnaire. Written summary instructions of procedures discussed in the training session were distributed to interviewers.

At each school approximately one-third of the interviewers simultaneously conducted interviews with all students from a particular class. Using this procedure, all pupils in kindergarten, 1st and 2nd grades were tested in each school. The testing was conducted individually and pupils were interviewed in a location on the school grounds that was chosen to ensure responses from other children would not be overheard.

Each testing session began with a brief set of instructions during which the subjects were assured of the confidentiality of their responses. As a component of these instructions, four example items were presented and discussed. Children were encouraged to indicate any difficulties they experienced in responding to an item. After reading each example item twice in rapid succession the interviewer asked the child if he/she understood the sentence. If the child did not understand the sentence the interviewer explained the sentence further, ascertained if the child understood the sentence, re-read the sentence, and requested a response. After ascertaining that the child understood the example item, the interviewer initially asked the child to respond "yes" or "no" to the sentence to indicate whether the sentence was true or false as a description of the child. If the child initially responded "yes" the interviewer then asked the child if he/she meant "yes always" or "yes sometimes". If the child initially responded "no" the interviewer then asked the child if he/she meant "no always" or "no

sometimes". The second response probe was stated for every response even when it was answered in the initial response (e.g. the child said "yes always" instead of "yes"), thus providing a check on the accuracy of the child's initial response. After the child successfully responded to example items and any questions were answered, the interviewers then read aloud each of the 64 positively worded SDQ I items. Halfway through the administration of the SDQI items the interviewer asked the child to do some physical activities for a brief period before proceeding to administer the remaining 32 items. This procedure was included to cater for young children with short attention spans. After presenting each of the first four items the interviewer asked the child if he/she understood the sentence before obtaining a response. The child was subsequently encouraged to indicate any difficulties he/she experienced in responding to the remainder of the items. This procedure was included to encourage children to seek clarification of any item they did not understand. If the child stated that the item was not understood the interviewer explained the meaning of the item further and ascertained the child understood the sentence before readministering the item. Children were periodically asked if they understood subsequent items during the remainder of the administration. Pilot work indicated that some kindergarten students had difficulty understanding a few of the items, and these items were initially presented in their original form and then paraphrased to ensure that they were understood. Thus, for example, children were told that mathematics meant work with numbers.

If a child did not initially respond to an item by stating yes or no, the interviewer explained the meaning of the sentence, re-read the sentence and requested a response. If the child still did not respond appropriately the item was circled and re-read after the administration of remaining items. If the child still did not respond appropriately the child was asked if he/she understood the sentence. If the child did not understand the item, the item was further clarified by the interviewer. If the child indicated he/she understood the sentence but could not decide whether to respond yes or no, the interviewer recorded a response of 2.5, halfway between the responses of "no sometimes" and "yes sometimes. Because the occurred infrequently and children were not told of this option, this middle category was used very infrequently.

Approximately 2 weeks after the individually administered SDQIs were collected, the SDQI was administered to nearly all the 2nd grade (n=158) children and a majority of the 1st grade (n=111) students using the typical group administration procedures (Marsh, 1988) instead of individually

administered interviews. The group administration procedure was deemed to be inappropriate for kindergarten children -- even after completion of the individualized administration procedure. Two classes of 1st grade students from one school were unable to participate in this second phase of the study for reasons unrelated to the purposes of the study. For purposes of the group administration, each child was given a copy of the SDQI. The researcher read the instructions aloud, clarified them, presented several example items, and then answered any questions before presenting the SDQI items. The SDQI items on the questionnaire given to each child were then read aloud twice and children were asked to respond to the items on their questionnaire. For purposes of the group administration, the standard five-point response scale (false, mostly false, sometimes true/sometimes false, mostly true, true) was used. The group administration procedures used here and those presented in the manual differed in that children were asked to place a ruler under the item being read aloud to facilitate marking their response on the correct line.

Statistical Analyses

The statistical analyses consisted of an evaluation of the psychometric properties (reliability and factor structure) of the self-concept responses and of sex and age effects in the self-concept ratings. Separate factor analyses were conducted on responses by each age group separately and for the total group, using the LISREL approach to CFA as in Marsh and Hocevar (1985). Multivariate and univariate ANOVA were used to test sex and age effects in the multiple dimensions of self-concept.

As part of the analyses, correlations among SDQI scores based on responses collected in the present investigation were compared with those based on responses in the normative archive of SDQI responses for students in grades 2-6 (Marsh, 1988; also see Marsh, 1989). Most SDQI research has used factor scores that are routinely produced by the SDQI scoring program, based on factor score coefficients derived from a factor analysis of all responses in the normative archive (Marsh, 1988, 1989). Whereas the scoring program computes both factor scores and simple scale scores based on an unweighted average of responses to the items designed to measure each scale, correlations among the SDQI scores are typically smaller for factor scores than for scale scores (Marsh, 1989). For purposes of the present investigation, correlations based on both sets of scores are considered. Because many of the responses in the normative archive are based on an earlier version of the SDQI that did not contain the General Self scale on the current version and considered here, the General Self scale was not included for purposes of just these comparisons with the normative archive.

Confirmatory factor analysis. As is typical in SDQ research (e.g., Marsh, 1988; Marsh & Hocevar, 1985) factor analyses were conducted on item-pair scores (or parcels) in which the first two items in each scale are averaged to form the first item pair, the next two items are used to form the second pair, and so forth. Thus the 64 SDQI items were reduced to 32 item pairs that were used in subsequent analyses. Analysis of item pairs instead of individual items is advantageous because the item pairs tend to be more reliable, to be more normally distributed, and to have less idiosyncratic variance than do individual items. Also, it is often recommended that there are at least 5 times as many subjects as variables in factor analyses and this guideline was satisfied for separate analyses at each grade level by factor analyzing item pairs instead of items.

In CFA, particularly when results from different samples are to be compared, it is appropriate to analyze covariance matrices instead of correlation matrices. Parameter estimates based on covariance matrices, however, are not so easily interpreted as the standardized measures based on correlation matrices. An appropriate compromise (Joreskog & Sorbom, 1988) is to standardize measured variables across the total sample and then to construct covariance matrices for each of the subsamples to be considered separately. This approach was used in the present investigation.

In CFA the researcher posits alternative a priori models to fit the data, compares the ability of the models to actually fit the data, and, perhaps, posits further a posteriori models if the a priori models do not adequately fit the data. For present purposes, three a priori models were fit to the data from each year group separately and to the total sample across all three year groups: (a) a one-factor model in which all measured variables loaded onto a single general factor factor; (b) a two-factor model in which the variables from the three academic scales defined an academic factor and the rest of the measured variables defined a non-academic factor; and (c) an 8-factor model in which each factor corresponded to one of the SDQI scales. In these models, each measured variable was allowed to load on only the one factor that it was posited to reflect (an independent clusters model). Factor correlations and uniquenesses (residuals or specific variances for each measured variable) were estimated, but correlations among the uniquenesses were restrained to be zero. Support for the a priori factor structure of the SDQI responses is based on the performance of the 8-factor model the corresponds to the design of the instrument.

An important unresolved issue is how to determine whether the fit is sufficient to justify support of an a priori model. In general the approach

is to evaluate the parameter estimates to determine whether they are consistent with predictions and to evaluate goodness of fit for alternative models. Researchers have developed a variety of goodness of fit indicators to aid in this process and those that appear to be among the most useful (see Bentler, in press; Marsh, Balla & McDonald, 1988; Marsh, McDonald & Balla, in press; McDonald & Marsh, in press) are considered here:

1. chi-square goodness of fit statistic (X^2) = N x FF; where N = sample size and FF is the maximum likelihood fitting function

2. Tucker-Lewis Index (TLI) = TLI = $(X_n^2/df_n - X_t^2/df_t)/(X^2_n/df_n - [1.0])$; where df=degrees of freedom and the subscripts n and t refer to values obtained from a null model and the target model respectively.

3. Unbiased Relative Fit Index (URFI) = $(Dk_n - Dk_t) / (Dk_n - 0)$; Note that this is an unbiased counter-part of the fit index originally proposed by Bentler and Bonett (1980; see Bentler, in press; McDonald & Marsh, in press) where Dk = FF - df/N.

Better fitting models have lower chi-squares and higher TLIs and URFIs. Although there are no clearly established rules as to what constitutes a "good" fit, a widely applied guideline for relative indices like the URFI and the TLI is .90 (e.g., Bentler & Bonett, 1980; Bentler, in press). An index of .90 can be roughly interpreted as being able to explain 90% of the covariation among the measured variables. The TLI and URFI differ primarily in that the TLI has a penalty for model complexity whereas the URFI does not. If none of the a priori models is able to fit the data adequately, the researcher may propose additional a posteriori models to better fit the data. LISREL provides modification indices (see Joreskog & Sorbom, 1983) that estimate the change in chi-square due to adding additional estimated parameters to the model. For example, the modification indices may suggest that a particular variable should be allowed to load on more than one factor even though the a priori model posited independent clusters.

Preliminary results — internal consistency estimates. Internal consistency estimates for the 8 individually administered scales (Table 1) are in the .70s and .80s for each year group and for the total sample except for the Parents (.692) and Physical (.505) scales with kindergarten respondents (see Table 1). In general, these reliability estimates increase with age (median estimates are .735, .797, and .819 for kindergarten, 1st-grade and 2nd grade students). For the three total scores, interestingly, the reliability estimates for the three age groups are more similar than are the estimates for individual scales, and those for 1st grade students are slightly higher than those of 2nd graders. The internal consistencies of the General self scale (.726, .781, .742, respectively) are moderate — though

below the median reliabilities for each group -- and show less age effects than do the averages of all scales. The internal consistencies for the group administered responses (administered to 1st and 2nd graders) show a similar pattern of results, though the size of the estimates is slightly higher. Overall, the internal consistency estimates provide reasonable support for the SDQI responses and indicate that responses to the General self scale are reasonably reliable for all three year groups. Whereas there are systematic age differences in the reliabilities of specific scales, age differences are smaller and less systematic for the three total scores and the General scale.

Results

Factor Structure For the Individually Administered SDQI Responses

Three a priori models were considered that posit (a) one general factor; (b) an academic and nonacademic factor, and (c) 8 factors corresponding to the 8 SDQI scales. These models were fit to the data from each year group separately and to the total sample across all three year groups. Support for the design of the SDQI requires requires that the 8-factor model is able to fit the data reasonably well, that it fits the data substantially better than the 1- and 2-factor models, and that parameter estimates based on the 8-factor model are reasonable.

Goodness of fit. Inspection of the URFIs (Table 2) indicates the 8-factor model fits the data substantially better than the 1-factor or 2-factor models. These results are consistent for the total group and for each year group considered separately. For the 1-factor and 2-factor models, the fit for the 2nd grade data is poorer than for either the kindergarten or 1st grade data. For the 8-factor model, however, the fit for the 2nd grade data is better than those for the younger children. These results indicate that the a priori 8-factor model consistent with the design of the SDQI does substantially better than models positing fewer factors, and that the advantage of the 8-factor model is larger for 2nd grade children. Using the £.90 guideline as a criterion of a "good" fit, the fit of the 8-factor a priori model is good for the total sample (.916), almost good for the 2nd grade data (.887) and the 1st grade data (.869) grade data, and somewhat less than good for the kindergarten data (.824).

Insert Table 2 About Here

Inspection of LISREL's modification indices (Joreskog & Sorbom, 1983) for the 8-factor model indicated that allowing measured variables to load on factors other than the one they were designed to measure would not improve fit substantially, but that freeing some correlations among uniquenesses associated with each measured variable would improve the fit. For the total

sample and each of the separate samples, the modification indices indicated that freeing correlated uniquenesses associated with the first two indicators of the Appearance factor and the first two indicators of the Reading factors would have a substantial effect. This suggests that the measured variables within each pair are more strongly correlated than can be explained by their relation to the common factor that they are designed to measure. Because of this consistency across the different samples, an 8-factor a posteriori model was tested in which these additional parameters were freed. Whereas the inclusion of these two additional parameters significantly improved the fit for the total sample and each subsample, the change in the URFIs was modest (Table 2). Because the conclusions based on the 8-factor a priori model differ little from those based on the corresponding a posteriori model, we will focus on the a priori model in subsequent discussion.

Parameter estimates. The evaluation of the factor loadings (see Table 3) for the 8-factor a priori models indicates that all 8 factors -- for each age group considered separately and for the total sample -- are well-defined; every factor loading is statistically significant and nearly all are substantial in size. The mean factor loading, however, is larger for older children.

Insert Table 3 About Here

Particularly for the youngest children we were concerned that the SDQI was so long that the quality of responses might deteriorate for items near the end of the questionnaire. Because the items within each scale are randomly ordered on the SDQI, inspection of the factor loadings from the first half of the SDQI with those from the second half provides a test of this possibility. Inspection of the factor loadings (Table 3), however, suggests that just the opposite occurred. The sizes of the factor loadings for the first two indicators are systematically lower -- not higher -- than those for the last two indicators, and the size of this difference is larger for the youngest children. These results suggest that -- particularly for the kindergarten children -- there was a practice effect such that the initial responses were systematically less effective than those of subsequent responses but that there was no apparent deterioration in responses near the end of the SDQI.

The size of factor correlations provide one indication of how well children are able to differentiate among the 8 factors. In evaluating the size of the factor correlations in CFA, it is important to note that these are correlations among latent constructs that have been corrected for unreliability and are thus larger than correlations between simple scale

scores [1] (see subsequent discussion of Tables 5 and 6). Nevertheless, the mean correlations among factors for both the kindergarten (.686) and 1st grade (.658) samples are substantial, whereas the mean correlation for the 2nd grade sample (.478) is substantially smaller. Despite the difference in mean correlations for different age groups, there is a consistent pattern in the relative sizes of the correlations. For all three samples the highest correlations involve the General and School scales. Specifically, the General Self scale is consistently correlated most highly with Physical Appearance, Peers, and the School scales, whereas the School scale is also highly correlated with the Reading and Maths scales.

The size of uniquenesses represents the specific variance and error variance in each measured variable that cannot be explained in terms of the 8 a priori factors. Consistent with results already discussed, the sizes of the uniquenesses decrease systematically with age (Table 3).[2]

The Harter Model. Harter and Pike (1984; Silon & Harter, 1985) used exploratory factor analyses to test the construct validity of responses to their self-concept instrument. Instead of the 4 scales that the instrument was designed to measure, only 2 factors were identified; a competence factor incorporating the physical and academic scales and an acceptance factor incorporating the social and maternal scales. In results summarized in this section we sought to evaluate this two-factor model using CFA instead of exploratory factor analysis.

The 4 scales the Harter and Pike instrument was designed to measure correspond most closely to the Peer Relationships, Parent Relationships, Physical Ability, and School scales from the SDQI.[3] We fit responses to just these SDQI scales with three different models; a one-factor model in which all variables loaded on one (general) factor, a two-factor model like that proposed by Harter, and a four-factor model in which each of the 4 scales defined a separate factor. For each age group and for the total sample, the Harter and Pike (two-factor) model performed marginally better than the one-factor model (see Table 4) but substantially poorer than the four-factor model. Consistent with earlier findings, the advantage of the four-factor model over the one- and two-factor models was positively related to the age of the children. These findings lead us to reject the two-factor model based on Harter and Pike (1984) and to conclude that -- even for young children -- self-concept is more differentiated than suggested by the Harter and Pike results.

Insert Table 4 About Here

Comparison of the Individually and Group Administered SDQIs

Students in just the 1st and 2nd grade samples completed the standard group administered SDQI approximately 2 weeks after the individually administered SDQIs. In the evaluation of the group administered responses several features are important. First, the group administration procedure is typically not recommended for children as young as these (Marsh, 1988), has only been used with 2nd grade students as part of one study (Marsh, Barnes, Cairns & Tidman, 1984), and has never been used with 1st grade children. Secondly, it is very likely that young children who — as in this study — have recently completed the individually administered SDQI will be better able to cope with the group administered SDQI than similar children who have not been previously exposed to the SDQI. Hence, an evaluation of the SDQI responses based on the group administration procedure in this study are unlikely to generalize to other samples. Given these considerations, several issues are relevant:

1. To the extent that children of a particular age are unable to cope with the group administered SDQI in the present investigation, other children of a similar age are even less likely to be able to do so in other studies in which the individually administered SDQI has not already been completed.

2. To the extent that the 8-factor model considered earlier is able to fit individually administered data better than the group administered data -- despite the likely advantage of group administered SDQI due to prior exposure to the individually administered SDQI — then there is strong support for the superiority of the individual administration.

3. To the extent that children respond appropriately to both the individually administered and group administered SDQIs, then the comparison of these responses provide useful information about the short-term stability of self-concept responses.

The most important comparison of the group-administered and individually-administered responses is the goodness of fits for the various models. For the group-administered responses, as with the individually administered responses, the 8-factor model is able to fit the data substantially better than the 1- and 2-factor models both for the separate groups and the total sample (Table 2). There are, however, important differences in the comparison of fits for the 1st and 2nd grade responses. Overall, fits are better for the individually administered data than for the group-administered data. Also, the differences between fits for 1st and 2nd grade data are larger for the group-administered data. For the individually-administered data, the 8-factor a priori model was able to fit the 1st grade data nearly as well as the 2nd grade data. In contrast, for the group-administered data, the fit for the 1st grade sample is substantially poorer

than for the 2nd grade data.

Parameter estimates for the individually- and group-administered responses are not directly comparable because of the different response scales. Several observations about the group-administered results (Table 3), however, are informative. The factor loadings for all 8 scales are statistically significant and substantial for both age groups and for the total sample. Factor correlations are very large (.682) for the 1st grade responses and substantially smaller for the 2nd grade sample (.425). The difference in mean correlations between the two year groups is also larger than observed for the individually administered responses (.671 and .478). The uniquenesses are also substantially larger for the 1st grade responses than for the second grade responses.

In summary, the 8-factor a priori model fits the individually administered responses better than the group administered responses. This difference is modest for the 2nd grade data, but more substantial for the 1st grade data. Because the design of the study was biased in favor of the group-administered responses -- since they came after the individually administered responses -- we interpret the results as demonstrating the superiority of the individually administered responses for both 1st and 2nd graders, but particularly for the 1st graders. Although the group administration procedure was not used with kindergarten students because it was deemed to be inappropriate, the advantages of the individual administrations can be assumed to be even larger for this age group.

Correlations Between Individually- and Group-Administered Scales

Results presented above suggest that the individual administration procedure is apparently effective with all three age groups whereas the group administration procedure is effective with 2nd graders and, perhaps to a lesser extent, 1st graders. Support for the group administration procedure, however, may not generalize to other studies in which this procedure does not follow the individual administration procedure. Nevertheless, at least for data in the present investigation, it is useful to examine correlations between scales derived from the two procedures. If both administration procedures were equally effective at inferring true self-concepts, then correlations between the two sets of scores would represent short-term stability. Because the two administration procedures are apparently not equally effective, the correlations reflect a combination of agreement between the two procedures and short term stability.

Scale scores for the two administration procedures were computed by taking the unweighted average of responses to the 8 items designed to

measure each scale. Correlations (Table 5) among the 16 scales — the 8 SDQI scales from each administration procedure — are presented separately for 1st-graders (above the main diagonal) and 2nd-graders (below the main diagonal). Each of these correlation matrices is a MTMM matrix in which the SDQI scales are the multiple traits, the two administration procedures are the multiple methods, and correlations between matching scales from the two administration procedures (those in < >) are convergent validities.[4] In evaluating MTMM matrices (e.g., Campbell & Fiske, 1957; Marsh, 1989), it is typical to consider convergent validity and discriminant validity separately. Convergent validity refers to the agreement between multiple methods of assessing the same trait and discriminant validity refers to the extent to which the traits are distinguishable.

Insert Table 5 About Here

Convergent validity. All the convergent validities for both year groups are statistically significant and most are substantial in size. The mean convergent validity is, however, substantially larger for the 2nd grade responses (.50) than for the 1st grade responses (.38). Correcting the convergent validities for unreliability (see Table 1) substantially increased the size of the coefficients, but did not reduce the difference in convergent validities for the 1st and 2nd grade responses (means of .47 and £.62, respectively).

Discriminant validity. Discriminant validity is typically assessed by comparing convergent validities (homotrait-heteromethod correlations) with correlations between different traits assessed by different methods (heterotrait-heteromethod coefficients) and with correlations between different traits assessed by the same methods (heterotrait-monomethod coefficients). Applying these two criteria:

1. For the 2nd grade data, convergent validities (mean = .50) are higher than heterotrait-heteromethod coefficients (mean = .23) for 99% of the 112 comparisons. For the 1st grade data, convergent validities (mean = 0.38) are higher than heterotrait-heteromethod coefficients (mean = .21) for 91% of the 112 comparisons.

2. For the 2nd grade data, convergent validities (mean = .50) are higher than heterotrait-homomethod coefficients (mean = .42) for 69% of the 112 comparisons. For the 1st grade data, convergent validities (mean = 0.38) are higher than heterotrait-homomethod coefficients (mean = .52) for only 28% of the 112 comparisons. For both age groups, heterotrait-homomethod coefficients for the group administered scales are higher and resulted in more violations of this criterion than did the those based on the individually administered scales.

The application of the traditional Campbell-Fiske criteria provide clear support for convergent validity and good support for at least one aspect of discriminant validity. Support for both convergent and discriminant validity was substantially stronger for the 2nd grade data than for the 1st grade data. There is, however, also evidence of a substantial method effect associated with each of the administration methods. This apparent method effect is larger for the 1st grade responses than the 2nd grade responses, and is larger for the group-administered responses than for the individually-administered responses.

Correlations Among SDQI Scores: A Comparison With the Normative Archive

The SDQI items used in both the individually and group administered procedures are the same as those used for students in grades 2 to 6 that comprise the normative archive for the SDQI test manual (Marsh, 1988; also see Marsh, 1989). Thus it is meaningful to compare correlations among SDQI scales in the present investigation with correlations based on the normative archive. Most SDQI research is based on factor scores instead of simple scale scores in part because correlations among the scores are consistently lower for factor scores (Marsh, 1988; Marsh, 1989; also see footnote 1). Comparisons based on both scale and factor scores are considered here. For present purposes, the mean of 28 correlations among 7 SDQI scores — all but General Self — and the mean of 7 factors selected a priori by Marsh (1989) to be among the lowest correlations were computed for each grade level (Table 6) [see Marsh, 1989, for further discussion of the rationale of these analyses and for findings based on high school student responses to the SDQII and late adolescent responses to the SDQIII].

Insert Table 6 About Here

Correlations among factor scores are consistently lower than correlations among scale scores, but the pattern of results is very consistent for both factor and scale scores. Results from the normative archive data (see Table 6) indicate that the mean correlations decline consistently with age at least through 5th grade and then appear to level out (also see Marsh, 1989). Also, the difference between the mean of correlations posited to be lower and the mean of all correlations is smaller for the youngest respondents. Marsh (1989) interpreted these results based on the normative archive to indicate that responses to the SDQI scales become more differentiated with age at least during the 2nd to 5th grade period.

A similar pattern of results is observed for the individually and group administered SDQI responses in the present investigation. For the individually administered responses, correlations among the SDQI scores for

2nd grade students are substantially smaller than those among 1st grade and kindergarten students. Because the 2nd grade responses are also more reliable, these differences would be even greater if the correlations were corrected for unreliability. Whereas correlations among 1st grade and kindergarten students are similar, the 1st grade responses are more reliable so that correlations corrected for unreliability are somewhat smaller for 1st grade students than for kindergarten students. This general pattern is also seen in responses to the group administered responses, though the difference between correlations based on 1st and 2nd grade responses is somewhat larger for the group administered responses than for the individually administered responses. Also, the correlations among 2nd grade responses are nearly the same for both group and individually administered responses, whereas correlations among 1st grade responses are somewhat larger for group administered responses than for individually administered responses.

Because both the normative archive and the present investigation include responses by 2nd grade students, it is interesting to compare the correlations based on these responses. The correlations among 2nd grade responses in the present investigation are substantially smaller than those in the normative archive. In fact, correlations based on 2nd grade responses in the present investigation are very similar to those based on 3rd grade responses in the normative archive. Although there are alternative interpretations of these findings, they suggest that 2nd grade students in the present investigation are better able to differentiate among the SDQI scales than 2nd grade students in the normative archive. We interpret this to mean that young children are apparently better able to cope with this the individually administered procedure than with the standard group administration procedures. The better differentiation based on the group administered responses in the present investigation apparently reflects the facilitative effect of already having completed the the individually administered task using the same SDQI items. These comparisons support earlier interpretations and suggest the superiority of the individually administered responses.

Sex and Age Effects

Although not a central focus of the present investigation, responses by children in the present investigation provide an opportunity to evaluate sex and age effects in self-concept for children who are younger than those typically considered. SDQI research (see Marsh, 1988, 1989) with slightly older children has consistently found that mean responses for most SDQI scales decline with age — Relations with Parents being a possible exception. In this previous research there has also been a consistent

pattern of counterbalancing sex differences that is apparently consistent with sex stereotypes. During preadolescent years the largest sex differences were for Physical Ability (favoring boys) and Reading (favoring girls). The observed sex differences were reasonably consistent across the early preadolescent to young adulthood period, with the apparent exception of Physical Appearance self-concept. In second grade girls had higher self-concepts of Physical Attractiveness than boys, but at older ages — particularly during high school years — girls had substantially lower self-concepts of Physical Appearance. These previous results provide a general basis of comparison for findings in the present investigation.

For purposes of the present investigation, a repeated measures ANOVA was used to assess the effects of age (kindergarten, 1st and 2nd grades) and sex across the 8 SDQI scales measured with the individually administered responses (see Marsh, 1989, for a more detailed overview of the analyses with older children). For total self-concept averaged across all 8 scales (i.e., the main effect of the repeated measure variable) the effects of sex, age, and their interaction were all nonsignificant. In each case, however, the sizes of these effects varied significantly depending on the SDQI scale. Tests of simple main effects (SPSS, 1986) were used to assess the effect of each scale (Table 7).

Insert Table 7 About Here

Age was significantly related to three SDQI scales: Physical Appearance, Peer Relations, and School. In each case only the linear effect of age was significant and the direction of the effects was negative. The effect of sex was statistically significant for three SDQI scales: Physical Ability, Physical Appearance, and Reading. Girls had substantially lower self-concepts of Physical Ability, and modestly higher self-concepts in Physical Appearance and Reading. There was also an age by sex interaction for Physical Ability. Whereas boys had higher self-concepts at all three ages, the sizes of the sex differences increased with age. The effects of sex and age were also assessed for each of the three SDQI total scores. The only effect to reach statistical significance, however, was the negative effect of age on the total nonacademic score.

In summary, the effects of sex and age were generally modest. Except for the large sex difference in self-concept of Physical Ability, none of the effects of sex, age, nor their interaction accounted for more than 2% of the variance in any of the SDQI scores. The direction of statistically significant effects — and even those that approached significance — were, however, similar to those found in other SDQI research with slightly older

children. In this respect, the consistent pattern of sex and age effects found here adds further support for the individually administered SDQI responses.

Discussion

The purposes of the present investigation were to evaluate a new, adaptive procedure for assessing multiple dimensions of self-concept for very young children and to seek answers to three theoretical questions. The central finding of the study is the clear support for the use of the individually administered SDQI for very young children. Due in part to the psychometric support for this new assessment procedure, the study was able to provide answers to three theoretical questions: (a) each of the 8 SDQI factors identified in responses by older children were identified here, indicating that self-concepts factor are better defined and more distinct for very young children than was previously assumed; (b) the general self-scale is apparently well-defined at each of the ages considered here, casting doubt on the suggestion that general self-concept does not evolve before 8 years of age; and (c) consistent with the Shavelson et al. model, the multiple dimensions of self-concept did appear to become more differentiated with age for these very young children. Each of these conclusions, however, warrants further consideration.

Very young children are much better able to differentiate among multiple dimensions of self-concept than previously assumed. Responses by older children to appropriately constructed self-concept instruments (e.g., Harter, 1982; Marsh, 1988) are well differentiated in that factor analyses have identified the factors that such instruments are designed to measure. Harter and Pike (1984; also see Silon and Harter, 1985), however, concluded that very young children are only able to differentiate between two facets of self-concept representing general competence and social acceptance. In contrast to the Harter and Pike conclusions, the 8 a priori SDQI factors found in responses by older children were identified for children aged 5 to 7. Even when we limited consideration to just the four self-concept scales most like those considered by Harter and Pike, there was clear evidence for the superiority of a four-factor model over the two-factor model proposed by Harter and Pike. The critical differences are apparently that we considered a wider variety of self-concept domains than did Harter and Pike and we employed CFA which is a substantially stronger analytic technique than the exploratory factor analysis procedures that they used. At least the second part of this suggestion could be tested by reanalyzing the Harter and Pike (1984) data using CFA instead of exploratory factor analyses.

General self-concept is reasonably well defined for the age range

considered in the present investigation. Harter (1983; Harter & Pike, 1984; Silon & Harter, 1985) reported that general self-concept responses did not emerge as a separate factor, were not reliable, and were not systematically related to the factors that were identified. In the present investigation our General self scale was reasonably reliable for each age group. CFA of responses for at each age group identified the General self factor and the factor loadings were substantial. Furthermore, the correlations among the General scale and the other scales were consistently among the largest of any or the correlations. Our conclusions clearly conflict with those of Harter, but so little information about the nature of their general self items and their analyses in the pilot study conducted by Harter and Pike is available that it is difficult to determine critical differences between the two studies.

The general self-concept scale in the Silon and Harter (1985) study was similar in design and -- based on responses by older children -- was substantially correlated with the General Self scale on the SDQI (Marsh & Gouvernet, 1989; Marsh & McDonald-Holmes, in press). The Silon and Harter study, however, was based on responses by educable mentally retarded children aged 9-12 who had mental ages of less than 8, apparently used the standard group administration procedures instead of individually administered instruments used in Harter and Pike, employed exploratory factor analyses instead of CFA, and did not actually report reliability estimates for their General scale nor correlations between this scale and other scales. A reanalysis of the Silon and Harter data that focused on their General scale and used CFA instead of exploratory factor analysis may resolve some of the apparent conflicts. Nevertheless, generalizations based on older, retarded children using a different instrument and different administration procedures to results based on younger, normal children should be interpreted cautiously.

Support for the existence of a reasonably well-defined General self-concept for very young children has important theoretical implications. Two apparently quite different models of the evolvement of self-concept have been proposed. Coopersmith (1967) and others have proposed that specific facets of self-concept evolve from a global sense of self. Harter (1983, 1986; Harter & Pike, 1984), however, claimed that general self-concept does not exist before the age of 8 and that this general self requires the young child to integrate the very concrete self-perceptions that young children have of themselves. Our results suggest that Harter's proposal may be wrong. Whereas Harter may be correct in her assumption that very young children do

not have the cognitive capacity to integrate systematically the self-concepts in specific domains, this may not be way they form a general self-concept. We find the existence of a general self-concept to be consistent with the "all-or-none" thinking that Harter (1983) has identified in very young children that apparently is not based on a systematic integration of domain-specific information. Relevant to this suggestion is the observation (Table 4) that the short-term stability of the General self scores are the lowest of all the SDQI scores for the 2nd grade students and particular for the 1st grade students. This suggests, perhaps, that the basis of General Self concept is more ephemeral than would be the case if it represented an integrated average of specific domains. It should be noted, however, that research with adolescents and even young adults (e.g., Marsh, in press-a; Marsh, Richards & Barnes, 1986a, 1986b) has also found General Self to be less stable than domain specific facets of self-concept.

The lack of support for the Harter's proposal may -- by implication -- support Coopersmith's alternative proposal. We feel, however, that this interpretation may be premature. Even though both domain specific and general self-concept factors were identified here, there is no basis for concluding that one preceded the other. Consistent with the Shavelson et al. model, a reciprocal pattern of relations in which General self-concept both affects and is affected by content-specific domains of self is also possible. Furthermore, the reasonably well-defined factor structure underlying SDQI responses is clearly inconsistent with the empirical basis of Coopersmith's proposal.

General self-concept for the very young children considered here was reasonably reliable at any particular time but was less stable over time than content specific dimensions of self. If General self were based on a systematic integration of content specific domains, however, it should logically be more stable. From this perspective it may not make sense to argue that distinctions among specific domains are made in relation to a pre-existing global sense of self. Instead, it appears that General self concept for very young children ages may reflect an unsystematic integration of specific domains of self concept that is easily swayed by mood or events of momentary salience. Furthermore, responses by older respondents to general self-scales may -- to a lesser extent -- also reflect such tendencies. For example, Marsh, Richards, and Barnes (1986a, 1986b) reported that responses by late-adolescents and young adults to the General self scale were the least stable over time of any of the 13 SDQIII scales even though it was one of the most reliable at each time considered separately.

Schwarz, Strack, Kommer, and Wagner (1987) described a particularly

relevant model of cognitive processing to explain why global judgments of subjective well being by adults were less stable than corresponding domain-specific judgments. According to their model, accurate global evaluations would "require time-consuming information processing, involving a systematic consideration of many aspects of one's life as well as a multitude of comparisons, and an integration of their implications into a single composite judgment" (p. 70). Because of the complexity of this task, they suggested that mood at the time is used as a judgmental short-cut or heuristic device for inferring subjective well being, and may also affect the availability of information in specific domains. In a test of their model, they found that minor events that impacted subject's mood state had more impact on global judgments than on domain-specific judgments. Although their focus was on global judgments of subjective well being, Schwarz (personal communication) found a similar — but weaker — pattern of results for Rosenberg-like measures of esteem. Thus, whereas adults apparently have the cognitive capabilities to more fully integrate domain-specific information in making global judgments, they apparently do not do so in most situations. In this respect, the cognitive basis of global judgments by adults may resemble those by young children found here.

Results of the present investigation support the Shavelson et al. (1976) hypothesis that self-concept becomes more differentiated with age. First, the average correlation among the SDQI factors becomes smaller with increasing age. Second, the difference in fits of models positing 1, 2, and 8 factors (or 1, 2 and 4 factors in tests of the Harter and Pike model) become larger with age. Finally, internal consistency estimates for the three SDQI total scores — total academic, total non-academic, and total self — do not vary substantially with age even though the 8 specific SDQI scales are substantially more reliable for older children. Also, comparisons with SDQI responses from the normative archive suggest that the individually administered, adaptive procedures may facilitate the differentiation of self-concept facets by very young children. Stipek and MacIver (1989) suggested that the failure of previous research to demonstrate the ability of very young children to differentiate among self-concept facets may be an artifact of difficulties introduced by existing self-concept measures, and our results support this suggestion. However, the support for the increasing differentiation of self-concept responses with age found here may also reflect differences in the ability of children to cope with even the individually administered SDQI. Whereas this possibility is always a viable alternative, it would apparently not be consistent with the finding that

responses by the youngest children were nearly as reliable as those of older children for the total scales and for the General self scale. Hence we cautiously interpret the results as offering support for the hypothesis that self-concept becomes more differentiated with age and that the identification of this differentiation is facilitated by the individually administered SDQI.

Factor analytic results presented here imply that very young children are cognitively able to differentiate among the content domains represented on the SDQI. This adds to a general trend demonstrating that children can perform a variety of cognitive tasks at younger ages than typically hypothesized. Whereas it is possible that the factors are merely statistical abstractions with no concrete reality or reflect response biases idiosyncratic to each scale, such counter-explanations seem implausible. In addition to the statistical support presented here, the spontaneous verbalizations of children in the interview setting suggested that they understood the difference between such content domains as physical appearance and school ability. However, alternative tests of this conclusion based on a different task (e.g., a sorting task based on the content of the items) could be used to test counter-explanations of our findings and, perhaps, to provide a developmental analysis of the processes through which self-concepts are formulated by very young children.

The Harter and Pike (1984) instrument was apparently the best available instrument for measuring multiple dimensions of self-concepts for very young children, but the results for the present investigation suggest that the psychometric properties of the individually administered SDQI are stronger. Because the construction of appropriate self-report instruments for young children is a pervasive problem, it is useful to speculate about why these differences exist. Harter and Pike (1984) presented children with parallel sets of (orally presented) verbal statements and pictures, whereas we used only verbal statements. It is plausible that the pictures would facilitate the task as suggested by Harter and Pike, but the need to process parallel stimulus inputs may have complicated the task. The four-point response scale consisting two dichotomous choices used by Harter and Pike (1984) is similar to the one used here, and so this is an unlikely basis of the difference. Both studies administered the materials individually, but the procedures used by Harter and Pike were not presented in sufficient detail to compare them with the procedures used here. Our procedures provided considerable opportunity for the administrators to clarify -- if necessary -- interpretations of the test items and the responses, but this may not have been the case in the Harter and Pike study. Because we measured twice as many factors as Harter and Pike, our instrument was considerably longer (64

items vs. 24 items). Whereas it is plausible that the shorter instrument would be more effective, our results showed an apparent practice effect such that items from the last half of the SDQI were more effective than items from the first half that was larger for younger children. Hence, the brevity of the Harter and Pike instrument may have been a weakness instead of a strength. Harter and Pike (1984) developed a new instrument specially for very young children whereas we adapted the existing SDQI for this purpose. Because the Harter and Pike instrument and its four a priori factors have not been validated with children at any age, the failure to support the a priori structure with very young children may reflect problems idiosyncratic to the instrument rather than general developmental trends. Also, comparisons between responses to their instrument and responses to different instruments by older children may differ because of the age of the children or the differences in the instruments. In contrast, the SDQI used here is well-validated with responses by slightly older children and the availability of this research facilitates the comparison of responses by very young children and by older children. Finally, the CFA used here was stronger than the exploratory factor analysis used by Harter and Pike, and we suspect that a reanalysis of their data with CFA would provide stronger support for their instrument as well as our conclusions about the factor structure of self in very young children. A detailed evaluation of these differences would require that both instruments were used in the same study and, perhaps, a systematic manipulation of differences in instrument construction. Hence, firm conclusions about why these differences exist are beyond the scope of the present investigation but warrant further consideration.

In summary, results of the present investigation provide support for a new procedure for assessing multiple dimensions of self-concept with very young children. Due in part to the success of this new procedure, we were able to provide new evidence for important issues related to the development of self-concept in very young children. In particular, the results show that self-concept is much better differentiated by very young children than has been previously assumed and that these children do have a global self-concept. The development of an improved procedure of assessing self-concept for very young children also has important theoretical and practical implications beyond those specifically considered here. The considerable advances in self-concept theory and practice for older children in the last decade was apparently based in part on advances in the ability to measure appropriately multiple dimensions of self-concept and the same may occur for research for very young children.

Footnotes

1 — Consistent with other research, studies based on all three SDQ instruments show that correlations among factors derived by CFA (e.g., Table 2) are larger than correlations among simple unweighted scale scores (e.g., Tables 5 and 6), which are higher than correlations among factor scores (e.g., Table 6), which in turn are higher than factor pattern correlations (not considered here) based on exploratory factor analyses such as those conducted with SPSSx (1986). Whereas a technical discussion of the basis for these differences is beyond the scope of the present investigation, it is important to note that the pattern of correlations is typically very similar for all the various sets of correlations.

2 — Subsequent tests of the invariance of parameter estimates across the three age groups similar to those described by Marsh and Hocevar (1985) indicated that there were significant differences between the groups. In order to reduce the complexity of the materials presented and because the nature of these differences are discussed in relation to parameter estimates presented for each group separately (see Tables 3 and 4), these subsequent tests of factorial invariance are not presented.

3 — This inference was based in part on studies (Marsh & Gouvernet, 1989; Marsh & MacDonald-Holmes, in press) that specifically compared the content of scales from the SDQI and the Harter (1982) instruments. and correlated responses from the two instruments using multitrait-multimethod analyses.

4 — The traditional MTMM term convergent validity is retained even though these correlations might be interpreted to reflect consistency or stability instead of validity. As noted by Marsh (1989) in his review of this analytic approach, MTMM analyses are appropriate when the different methods are very similar or very dissimilar.

REFERENCES

Bentler, P. M. (in press). Comparative fit indices in structural models. Psychological Bulletin, .

Bentler, P. M. & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletih, 88, 588-606.

Byrne, B. M. (1984). The general/academic self-concept nomological network: A review of construct validation research. Review of Educational Research, 54, 427-456.

Byrne, B. M. (in press). A review of methodological approaches to the validation of academic self-concept: The construct and its measures.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56., 81-105.

Coopersmith, S. A. (1967). The antecedents of self-esteem. San Francisco: W. H. Freeman.

Dusek, J. B., & Flaherty, J. F. (1981). The development of self-concept during adolescent years. Monographs of the Society for Research in Child Development, 46, (4,Serial No. 191).

Fleming, J. S., & Courtney, B. E. (1984). The dimensionality of self-esteem: II: Hierarchical facet model for revised measurement scales. Journal of Personality and Social Psychology, 46, 404-421.

Harter, S. (1982). The Perceived Competence Scale for Children, Child Development, 53, 87-97.

Harter, S. (1983). Developmental perspectives on the self-system. In P. H. Mussen (Ed.), Handbook of Child Psychology, (Volume IV, 4th edition, pp. 275-385). New York: Wiley.

Harter, S. (1985). Processes underlying the construction, maintenance, and enhancement of the self-concept in children. In J. Suls & A. G. Greenwald (Ed.),The development of self (pp. 137-181). Hillsdale, NJ: Lawrence Erlbaum.

Harter, S. (1986). Processes underlying the construction, maintenance, and enhancement of self-concept in children. In S. Suls & A. Greenwald (Ed.), Psychological perspectives of the self (Vol. 3), (pp. 136-182). Hillsdale, NJ: Erlbaum.

Harter, S. (1988). The construction and conservation of self: James and Cooley revisited. In D. K. Lapsley & F. C. Power (Ed.), Self, ego and identity: Integrative approaches (pp. 43-70). New York: Springer-Verlag.

Harter, S., & Pike, R. G. (1981). The pictorial perceived competence scale and acceptance for young children. Unpublished manuscript, University of

Denver.

Harter, S., & Pike, R. (1984). The pictorial scale of perceived competence and social acceptance for young children. Child Development, 55, 1969-1982.

James, W. (1890/1963). The principles of psychology. New York: Holt, Rinehart & Winston.

Joreskog, K. G. & Sorbom, D. (1981). LISREL V: Analysis of Linear Structural Relations By the Method of Maximum Likelihood. Chicago: International Educational Services.

Joreskog, K. G., & Sorbom, D. (1988). LISREL 7: A guide to the program and applications. Chicago: SPSS, Inc.

Markus, H., & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. Annual Review of Psychology, 38, 299-337.

Marsh, H. W. (1986a). The bias of negatively worded items in rating scales for young children: A cognitive-developmental phenomena. Developmental Psychology, 22, 37-49.

Marsh, H. W. (1986b). Global self esteem: Its relation to specific facets of self-concept and their importance. Journal of Personality and Social Psychology, 51, 1224-1236.

Marsh, H. W. (1987). The hierarchical structure of self-concept and the application of hierarchical confirmatory factor analysis. Journal of Educational Measurement, 24, 17-19.

Marsh, H. W. (1988). Self Description Questionnaire: A Theoretical and empirical basis for the Measurement of multiple dimensions of preadolescent self-concept: A test manual and a research monograph. San Antonio, TX: The Psychological Corporation.

Marsh, H. W. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to Early-adulthood. Journal of Educational Psychology..

Marsh, H. W. (In Press-a). A multidimensional, hierarchical model of self-concept: Theoretical and empirical justification. Educational Psychology Review.

Marsh, H. W. (in press-b). Multitrait-multimethod analyses. In J. P. Keeves (Ed.), Educational research methodology, measurement and evaluation: An international handbook. Oxford, Pergamon Press.

Marsh, H. W., Balla, J. R. & McDonald, R. P. (1988). Goodness-of-fit indices in confirmatory factor analysis: The effect of sample size. Psychological Bulletin, 102, 391-410..

Marsh, H. W., Barnes, J., Cairns, L., & Tidman, M. (1984). The Self Description Questionnaire (SDQ): Age effects in the structure and level of

self-concept for preadolescent children. Journal of Educational Psychology, 76, 940-956.

Marsh, H. W., Byrne, B. M., & Shavelson, R. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. Journal of Educational Psychology, 80, 366-380.

Marsh, H. W., & Gouvernet, P. (1989). Multidimensional self-concepts and perceptions of control. Construct validation of responses by children. Journal of Educational Psychology, 81, 57-69.

Marsh, H. W., & Hocevar, D. (1985). The application of confirmatory factor analysis to the study of self-concept: First and higher order factor structures and their invariance across age groups. Psychological Bulletin, 97, 562-582.

Marsh, H. W. & MacDonald-Holmes, I. W. (In Press). Multidimensional self-concepts: Construct validation of responses by children. American Education Research Journal, .

Marsh, H. W., McDonald, R. P., & Balla, J. R. (in press). Goodness-of-fit indices in confirmatory factor analysis: The effect of sample size and model parsimony. Multivariate Behavioral Research, .

Marsh, H. W., & Richards, G. (1988) The Outward Bound Bridging Course for low achieving high-school males: Effect on academic achievement and multidimensional self-concepts. Australian Journal of Psychology, 40, 281-298.

Marsh, H. W., Richards, G., & Barnes, J. (1986a). Multidimensional self-concepts: A longterm followup of the effect of participation in an Outward Bound program. Personality and Social Psychology Bulletin, 12, 475-492.

Marsh, H. W., Richards, G., & Barnes, J. (1986b). Multidimensional self-concepts: The effect of participation in an Outward Bound program. Journal of Personality and Social Psychology, 45, 173-187.

Marsh, H. W., & Shavelson, R. J. (1985). Self-concept: Its multifaceted, hierarchical structure. Educational Psychologist, 20 107-125.

Marsh, H. W., & Smith, I. D. (1982). Multitrait-multimethod analyses of two self-concept instruments. Journal of Educational Psychology, 74, 430-440.

Marx, R. W., & Winne, P. H. (1978). Construct interpretations of three self-concept inventories. American Educational Research Journal, 15, 99-108.

McDonald, R. P. (1985). Factor analysis and related methods. Hillsdale, NJ: Erlbaum.

McDonald, R. P, & Marsh, H. W. (in press). Choosing a multivariate model: Noncentrality and goodness-of-fit. Psychological Bulletin, .

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (in press). Evaluation of goodness-of-fit indices for

structural equation models. Psychological Bulletin, 105, 430-445.

Rosenberg, M. (1965). Society and the adolescent child. Princeton: Princeton University Press.

Rosenberg, M. (1979). Conceiving the self. New York: Basic Books.

Schwarz, N., Strack, F., Kommer, D., & Wagner, D. (1987). Soccer, rooms, and the quality of your life: Mood effects on judgments of satisfaction with life in general and with specific domains. European Journal of Social Psychology, 17, 69-79.

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Validation of construct interpretations. Review of Educational Research, 46, 407-441.

Silon, E. L., & Harter, S. (1985). Assessment of perceived competence, motivational orientation, and anxiety in segregated and mainstreamed educable mentally retarded children. Journal of Educational Psychology, 77, 217-230.

SPSS (1986). SPSSx User's Guide. New York: McGraw-Hill.

Stipek, D. J. (1981). Children's perceptions of their own and their classmates' ability. Journal of Educational Psychology, 73, 404-410.

Stipek, D. J., & Mac Iver, D. (1989). Developmental change in children's assessment of intellectual competence. Child Development, 60, 521-538.

Stipek, D. J., & Weisz, J. R. (1981). Perceived personal control and academic achievement. Review of Educational Research, 51, 101-137.

Winne, P.H., Marx, R. W., & Taylor, T. D. (1977). A multitrait-multimethod study of three self-concept inventories. Child Development, 48, 893-901.

Wylie, R. C. (1974). The self-concept. (Rev. ed., Vol. 1) Lincoln: University of Nebraska Press.

Wylie, R. C. (1979). The self-concept. (Vol. 2) Lincoln: University of Nebraska Press.

Wylie, R. C. (in press). Measures of self-concept. Lincoln: University of Nebraska Press.

Table 1

Coefficient Alpha Estimates of Reliability for Each Grade Level and the Total Sample: Individual (Ind) and Group (Grp) Administrations

	Kindergarten		1st Grade		2nd Grade		Total Sample	
Scale	Ind	Grp	Ind	Grp	Ind	Grp	Ind	Grp
Physical	.505	...	.710	.782	.730	.745	.668	.764
Appearance	.744	...	.830	.832	.861	.797	.826	.814
Peers	.770	...	.753	.809	.807	.842	.786	.828
Parents	.692	...	.726	.811	.765	.837	.722	.825
Read	.757	...	.841	.866	.837	.827	.820	.841
Math	.773	...	.833	.846	.853	.866	.823	.856
School	.724	...	.812	.786	.831	.868	.796	.839
General	.726	...	.781	.818	.742	.782	.749	.799
Total Scores								
Non-Academic	.845	...	.885	.916	.879	.886	.870	.902
Academic	.890	...	.910	.917	.902	.920	.903	.918
Total	.929	...	.947	.956	.939	.944	.939	.950

Note. The individually administered SDQIs were obtained from kindergarten, 1st and 2nd grade students, whereas the group administered SDQI were obtained from only 1st and 2nd grade students. Coefficient alpha estimates of reliability depend on the mean correlation among items and the number of items. Hence, the total scores — which are based on more items than the scale scores — have higher reliabilities even though the mean of correlations among items tends to be smaller.

Table 2

Goodness of Fit Indices for Alternative Models for Kindergarten (K), 1ST Grade, 2nd Grade, and the Total Sample (TOT) For Responses to the Individually and Group Administered Responses

		Individual			Group		
MODEL	df	Chi-Sq	TLI	URFI	Chi-Sq	TLI	URFI
Null Models[a]							
K	496	2399.59	0	0	—	—	—
1ST	496	2976.20	0	0	2743.26	0	0
2ND	496	3086.68	0	0	3320.97	0	0
TOT	496	6746.06	0	0	5159.94	0	0
1 Factor Models							
K	464	1058.75	.666	.688	—	—	—
1ST	464	1191.28	.687	.707	1397.29	.556	.585
2ND	464	1595.39	.533	.563	1770.20	.506	.538
TOT	464	2270.48	.691	.711	2352.66	.567	.595
2 Factor Models							
K	463	947.68	.727	.745	—	—	—
1ST	463	1109.76	.721	.739	1299.41	.601	.628
2ND	463	1463.84	.586	.614	1478.79	.615	.640
TOT	463	1942.99	.746	.763	1948.10	.659	.682
8 Factor Models (a priori)							
K	436	771.30	.800	.824	—	—	—
1ST	436	760.33	.851	.869	1015.20	.707	.742
2ND	436	729.38	.871	.887	884.58	.819	.841
TOT	436	961.49	.904	.916	1166.41	.822	.843
8 Factor Models (a posteriori)							
K	434	735.64	.819	.842	—	—	—
1ST	434	721.18	.868	.884	957.46	.734	.767
2ND	434	695.64	.885	.899	866.85	.825	.847
TOT	434	826.85	.928	.937	1094.73	.838	.858

Note. Chi-Sq = chi-square, TLI = Tucker-Lewis Index, URFI = unbiased relative fit index. Because the same model was fit to both individually and group administered data, the df are the same for both sets of data.

a — The null model posits that all of the measured variables are uncorrelated, and is used to define the (poorest fitting) endpoint for the TLI and the URFI.

Table 3

Summary of Parameter Estimates for Confirmatory Factor Analyses of Three Age Groups and the Total Group for the 8-factor A Priori Model

	Individual Administration Factor Loadings				Group Administration Factor Loadings		
Measured Variables	Kinder garten	1st Grade	2nd Grade	Total Group	1st Grade	2nd Grade	Total Group
Appearance							
1	.211	.592	.666	.507	.590	.588	.601
2	.394	.543	.556	.503	.740	.607	.649
3	.625	.642	.889	.695	.828	.744	.796
4	.550	.709	.601	.656	.698	.580	.634
Physical							
1	.432	.539	.937	.647	.703	.559	.648
2	.478	.643	.931	.696	.669	.723	.722
3	.716	.786	.578	.723	.817	.579	.681
4	.642	.828	.623	.724	.773	.723	.728
Peers							
1	.454	.564	.721	.584	.411	.664	.564
2	.752	.599	.696	.702	.654	.788	.727
3	.561	.606	.744	.647	.776	.705	.734
4	.876	.679	.704	.750	.876	.709	.776
Parents							
1	.550	.568	.584	.568	.670	.503	.572
2	.522	.500	.434	.478	.681	.554	.623
3	.748	.743	.715	.726	.771	.853	.800
4	.675	.813	.738	.758	.824	.811	.813
Read							
1	.589	.718	.652	.659	.875	.598	.717
2	.653	.775	.808	.751	.801	.679	.717
3	.743	.717	.803	.761	.773	.882	.826
4	.741	.770	.785	.784	.884	.774	.832
Math							
1	.408	.562	.697	.570	.669	.709	.697
2	.747	.650	.699	.701	.855	.808	.828
3	.766	.836	.868	.812	.767	.921	.858
4	.749	.696	.898	.794	.747	.884	.819
School							
1	.344	.619	.783	.602	.546	.669	.649
2	.580	.713	.626	.653	.638	.743	.703
3	.778	.701	.639	.705	.617	.807	.746
4	.698	.626	.960	.768	.732	.901	.821
General							
1	.482	.531	.467	.520	.860	.543	.669
2	.673	.679	.659	.660	.680	.652	.665
3	.783	.771	.702	.744	.778	.748	.770
4	.610	.757	.682	.679	.709	.690	.717
Mean Factor Loading	.610	.671	.714	.663	.732	.709	.722
Mean Factor Correlation	.686	.658	.478	.596	.682	.425	.579
Mean measured variable Uniqueness	.643	.499	.448	.540	.511	.429	.473

Note. For the 8-factor a priori model, each factor is inferred on the basis of four measured variables. Each measured variable was allowed to load on only the factor that it was designed to measure and all other factor loadings were constrained to be zero. For this reason, the factor loadings from each analysis are presented as a single column even though they represent 8 different factors. All factor loadings are statistically significant (standard errors typically vary between .06 and .09). Whereas only the means of factor correlations and measured variable uniquenesses are presented, the pattern of correlations in each analysis are generally similar to those among scales scores presented in Table 5 (though correlations among scales scores are not corrected for unreliability and thus tend to be smaller).

Table 4

Goodness of Fit Indices for Alternative Models for Kindergarten (K), 1ST Grade, 2nd Grade, and the Total Sample (TOT) Based on the Harter and Pike (1984) Model

	Goodness of Fit Indicators			
MODEL	Chi-Sq[a]	df	TLI	URFI
Null Models				
K	761.95	120	0	0
1ST	970.08	120	0	0
2ND	1106.76	120	0	0
TOT	2401.15	120	0	0
1 Factor Models				
K	232.88	104	.768	.799
1ST	259.29	104	.789	.817
2ND	502.30	104	.534	.596
TOT	655.40	104	.721	.758
2 Factor Models				
K	211.17	103	.804	.831
1ST	207.81	103	.856	.877
2ND	428.08	103	.616	.671
TOT	528.72	103	.783	.813
4 Factor Models				
K	164.67	98	.873	.896
1ST	147.15	98	.929	.942
2ND	161.47	98	.921	.936
TOT	215.39	98	.937	.949

Note. Chi-Sq = chi-square, TLI = Tucker-Lewis Index, URFI = unbiased relative fit index. The 4 SDQI factors (Physical Ability, Peers, Parents, School) that most closely match those proposed by Harter and Pike (1984) are considered in models summarized here. The two factor model corresponds to the model proposed by Harter and Pike in which one factor (competence) incorporates the Physical Ability and School scales, whereas the second (acceptance) incorporates the Peers and Parents scales.

a -- The null model posits that all of the measured variables are uncorrelated, and is used to define the (poorest fitting) endpoint for many of the goodness of fit indices.

Table 5

Multitrait-Multimethod Matrices of Relations Between Individually and Group Administered Self-Concept Responses For 1st Graders (above the main diagonal) and 2nd Graders (below the main diagonal)

	Individual											Group										
	Scales								Total Scores			Scales								Total Scores		
Phys	---	.42	.57	.39	.46	.43	.62	.56	.76	.62	.74	(.40)	.24	.23	.02	.30	-.02	.21	.17	.27	.19	.25
Appr	.27	---	.49	.43	.36	.41	.50	.62	.78	.52	.69	.14	(.49)	.34	.22	.15	.15	.18	.15	.37	.19	.30
Peer	.34	.36	---	.55	.41	.43	.57	.61	.83	.57	.75	.15	.30	(.44)	.21	.24	.13	.16	.19	.34	.20	.29
Prnt	.31	.48	.46	---	.36	.46	.49	.60	.74	.53	.68	.14	.21	.30	(.36)	.13	.14	.11	.15	.31	.15	.25
Read	.29	.42	.49	.43	---	.35	.59	.49	.51	.78	.71	.14	.16	.12	.12	(.40)	.05	.27	.05	.16	.28	.24
Math	.19	.27	.42	.40	.37	---	.57	.59	.55	.79	.74	.20	.29	.25	.28	.41	(.41)	.35	.22	.31	.45	.42
Schl	.26	.43	.43	.48	.59	.47	---	.72	.70	.87	.86	.27	.30	.31	.23	.44	.19	(.34)	.16	.34	.37	.39
Genl	.36	.57	.53	.54	.45	.49	.54	---	.77	.74	.81	.20	.37	.37	.22	.34	.18	.31	(.19)	.36	.32	.37
TNACD	.63	.76	.75	.76	.56	.44	.55	.70	---	.72	.92	.26	.41	.42	.26	.26	.13	.21	.21	(.42)	.23	.35
TACD	.30	.46	.55	.54	.82	.76	.85	.61	.64	---	.94	.25	.31	.28	.26	.51	.27	.39	.18	.34	(.45)	.43
TSELF	.50	.66	.70	.70	.78	.68	.79	.72	.88	.93	---	.28	.38	.37	.28	.43	.22	.33	.21	.41	.38	(.42)
Phys	(.53)	.09	.13	.11	.18	.02	.09	.17	.28	.12	.21	---	.41	.51	.47	.51	.39	.53	.49	.72	.55	.68
Appr	.20	(.45)	.19	.17	.30	.14	.25	.35	.36	.29	.35	.38	---	.64	.52	.38	.47	.37	.57	.80	.47	.68
Peer	.30	.31	(.54)	.27	.35	.10	.23	.35	.49	.28	.42	.35	.51	---	.72	.47	.61	.56	.69	.88	.64	.82
Prnt	.09	.24	.17	(.47)	.24	.20	.28	.26	.33	.30	.34	.17	.36	.37	---	.47	.63	.59	.66	.83	.65	.80
Read	.21	.21	.23	.24	(.61)	.19	.36	.23	.30	.49	.45	.31	.38	.39	.46	---	.46	.77	.44	.56	.86	.77
Math	-.01	.10	.17	.17	.20	(.55)	.28	.22	.15	.42	.33	.05	.25	.31	.50	.42	---	.64	.64	.65	.81	.79
Schl	.16	.19	.24	.30	.44	.33	(.44)	.25	.33	.50	.47	.26	.44	.45	.57	.72	.60	---	.60	.63	.92	.84
Genl	.22	.35	.39	.33	.44	.24	.34	(.41)	.45	.42	.48	.42	.65	.71	.57	.54	.37	.56	---	.74	.65	.75
TNACD	.38	.38	.37	.36	.38	.16	.30	.39	(.51)	.35	.47	.64	.78	.79	.67	.53	.40	.60	.82	---	.71	.92
TACD	.14	.19	.25	.31	.50	.42	.43	.28	.31	(.55)	.49	.24	.42	.45	.60	.85	.78	.91	.58	.60	---	.93
TSELF	.28	.31	.34	.37	.50	.34	.41	.37	.44	.52	(.53)	.46	.64	.67	.70	.79	.68	.87	.76	.87	.92	---

Convergent Validities
Corrected For Unreliability

1st (.54)(.59)(.56)(.47)(.47)(.49)(.43)(.23) (.47)(.49)(.44)

2nd (.72)(.55)(.66)(.59)(.73)(.65)(.52)(.54) (.58)(.60)(.56)

Note. Convergent validities, the values in ( ), refer to agreement between matching SDQI scales from the individually administered and group administered scales. Convergent validities were also corrected for unreliability (using reliability estimates from Table 1).

Table 6

Summary of Scale Distinctiveness Analyses For Responses to the SDQI

Sample and Age Level	N	Mean α	Mean Correlation Among: All Scales: Scale Scores	All Scales: Factor Scores	Selected Scales: Scale Scores	Selected Scales: Factor Scores
SDQI (Normative Archive)						
Grade 2	176	.83	.55	.43	.49	.37
Grade 3	107	.76	.37	.27	.30	.20
Grade 4	513	.86	.34	.24	.23	.12
Grade 5	1,428	.86	.27	.18	.18	.08
Grade 6	1,111	.87	.28	.18	.17	.07
SDQI (Individually Administered)						
Grade K	164	.71	.45	.35	.47	.37
Grade 1	169	.79	.48	.36	.48	.35
Grade 2	169	.81	.38	.28	.30	.19
SDQI (Group Administered)						
Grade 1	113	.82	.53	.43	.46	.33
Grade 2	158	.83	.38	.29	.29	.19

Note. Correlations among SDQI scales were computed for each grade level in the normative archive (Marsh, 1988; also see Marsh, 1989), for the individually administered responses, and the group administered responses. The means of the 28 correlations among all 7 SDQI scales (excluding the General scale for purposes of this analysis) and 7 correlations selected a priori by Marsh (1989) to be the lowest correlations were computed. Separate sets of correlations were computed for the simple scale scores (an unweighted average of responses designed to measure each scale) and for the factor scores (a weighted average of responses based on the a factor analysis that is part of the scoring program described by Marsh, 1988).

Table 7

Sex and Age Effects in the SDQI Scale and Total Scores

SDQ Score	Sex	Year in School			Effect Size and Statistical Significance		
SDQI Scales		Kinder-garten	1st Grade	2nd Grade	Age Linear	Sex	Sex by Age Linear
Physical	Boys	52.70	53.09	54.18	-.03	-.34**	-.10*
	Girls	47.37	48.06	44.10			
Appearance	Boys	50.77	49.37	47.54	-.18**	.08*	-.05
	Girls	53.50	51.00	47.99			
Peers	Boys	50.77	50.16	48.63	-.08*	.02	.01
	Girls	50.62	51.36	48.93			
Parents	Boys	48.02	48.66	51.30	.07	.06	-.06
	Girls	50.62	50.21	51.03			
Read	Boys	48.88	50.64	48.42	.01	.08*	.03
	Girls	50.48	50.76	51.47			
Math	Boys	51.52	51.07	48.97	-.03	-.04	.07
	Girls	48.70	50.45	49.61			
School	Boys	51.11	51.44	47.13	-.11**	.02	.05
	Girls	50.87	50.32	49.50			
General	Boys	49.22	50.22	49.13	-.04	.05	-.03
	Girls	51.15	50.81	49.54			
Total Scores							
Non-Academic	Boys	50.74	50.33	50.29	-.08*	-.05	-.06
	Girls	50.94	50.33	47.40			
Academic	Boys	50.59	51.24	47.85	-.05	.02	.06
	Girls	50.01	50.61	50.24			
Total Self	Boys	50.72	50.91	48.84	-.07	.01	.01
	Girls	50.46	50.53	48.88			

Note. All SDQI scores were standardized to have Mean = 50 and SD = 10 across the total sample. The effects of the quadratic component of age and its interaction with sex were also tested, but they were excluded because they were not significantly related to any of the self-concept scores. Effect sizes are standardized beta weights.

* $p < .05$; ** $p < .01$.